R E P O R T R E S U M E S

ED 011 146 EA 000 382

CRITERIA FOR METHODOLOGICAL ADEQUACY FOR RESEARCH ON EDUCATIONAL CHANGE.

BY- GEPHART, WILLIAM J.

PUB DATE SEP 65

EDRS PRICE MF-$0.18 HC-$2.40 60P.

DESCRIPTORS- *EDUCATIONAL CHANGE, *EDUCATIONAL RESEARCH, TECHNIQUES, *RESEARCH METHODOLOGY, MODELS, EVALUATION, RESEARCH, MILWAUKEE

RESEARCH ADEQUACY MUST BE ASSESSED AND STANDARDS DRAWN IF PROGRESS IS TO BE MADE IN THE ACCUMULATION OF KNOWLEDGE. THIS DISCUSSION OF METHODOLOGICAL CRITERIA FOCUSES UPON THE FOLLOWING TOPICS--(1) A LOGIC FRAMEWORK FOR EDUCATIONAL RESEARCH, (2) GENERAL CRITERIA FOR RESEARCH EVALUATION, (3) ELEMENTS OF THE STUDY OF THE EDUCATIONAL CHANGE PROCESS, (4) METHODS AND TECHNIQUES FOR STUDYING THE CHANGE PROCESS COMPONENTS, AND (5) CRITERIA OF ADEQUACY FOR EVALUATING RESEARCH TECHNIQUES IN THE STUDY OF EDUCATIONAL CHANGE. (GB)

ED011146

CRITERIA FOR METHODOLOGICAL ADEQUACY FOR RESEARCH ON EDUCATIONAL CHANGE

By

William J. Gephart
Director of Research and Experimentation
University of Wisconsin-Milwaukee

September 1965

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EA 000 382

CRITERIA FOR METHODOLOGICAL ADEQUACY FOR RESEARCH ON EDUCATIONAL CHANGE

The development of criteria for methodological adequacy of educational research has been a problem faced by professional educators for almost fifty years. In any absolute sense it is and should remain an unsolved problem as the field anticipates the evolution of improvements in research strategies, methodologies, and techniques. Thus, a statement of criteria for evaluating research is of value only within specific temporal boundaries, yet at the same time specification of evaluative criteria is highly necessary. Unless a systematic assessment of the strengths and weaknesses of existing research on the educational change process is undertaken, two deterrents to progress exist. First, since individual research efforts vary in adequacy, "facts" generated by these studies vary in value. Second, the development of improved strategies, methods, and techniques for research on educational change rests heavily on the analysis of existing techniques. Both of these are stumbling blocks to the continued accumulation of knowedge necessary for "growing" the "inductive inference tree" described by Platt as crucial to advancement in a substantive area.[1]

---

[1] J. R. Platt, "Strong Inference," Science. 146:347-52; October, 1964.

The evaluative criteria presented in this paper have evolved from two sources, literature and research on the research process. Much of the literature on the research process exists in the form of textual materials which contain the rationale and elaboration for the evaluative criteria presented here. There is also an expanding body of research literature in which research is the substantive topic. Six discrete directions can be observed in this literature.

1. The identification of type and frequency of errors found in educational research.[2]

2. The assessment of the content and form of research reports.[3]

3. The assessment of the value of research through study of its impact on textual materials.[4]

4. The identification of type and frequency of inadequacies in research proposals.[5]

---

[2]F. L. Whitney, The Elements of Research: Revised Edition. New York: Prentice-Hall, Inc., 1942. p. 55-7. Whitney presents in tabular form lists by eight authors in which research errors in attitude, method, and technique are identified. The summarized papers span a period from 1919 to 1930.

[3]G. M. Wilson, "Research: Suggested Standards for Summarizing and Reporting Applied to Two Recent Summaries of Studies in Arithmetic." Journal of Educational Research. 28:187-94, November 1934.

[4]Mother M. C. Dooley, 'The Relationship Between Arithmetic Research and the Content of Arithmetic Textbooks (1900-1957)," The Arithmetic Teacher 7:178-83; April 1960.

[5]G. R. Smith, "Inadequacies in a Selected Sample of Research Proposals." Unpublisned doctoral dissertation, Teachers College, Columbia University, 1964.

5. The analysis of the role of theory in the literature on the research process.[6]

6. The identification of the techniques, methods, or designs employed in educational research.[7]

Since the latter type of study--the use of judges to evaluate the adequacy of research--relates most directly to the current project, an expanded discussion is presented. Two general approaches to the use of expert judges in evaluating research adequacy seem to be employed. One group of studies involves the identification of one or a number of eminently qualified persons and asking them to evaluate selected research. The second approach also involves the selection of qualified persons but asks them to employ some specified evaluative criteria. Examples of the product of the unstructured approach can be seen in the research evaluation contained in the Review of Educational Research, a study of research on counseling and guidance,[8] a study of research in teacher education.[9] The structured approach is illustrated in

---

[6]K. E. Lake, "Inductive Methodology Versus Hypothetic-Deductive Methodology in Educational Research." Unpublished doctoral dissertation, University of Kansas, 1961.

[7]H. H Bixler, "Check Lists for Educational Research," New York: Teachers College, Columbia University, 1928. p. 85-7.

[8]W. B. King, Survey of the Status of Research in Guidance and Counseling. Washington, D.C.: U.S.O.E. Cooperative Research Project Number F-1, 1962.

[9]F. R. Cyphert and E. Spaights, An Analysis and Projection of Research in Teacher Education, Washington, D.C.: U.S.O.E. Cooperative Research Project Number F-015, 1964.

the work of Johnson,[10] the American Institute of Research,[11] and Gephart.[12]

The current project attempts the synthesis of these two approaches, as a statement of criteria for evaluating research on educational change has been employed by the project staff, and further unstructured evaluation by persons selected for specific competencies is scheduled during the conference.

The discussion which follows will focus on the criteria of methodological adequacy. To do so, it will treat sequentially the following topics:

1. A plausible logic framework for educational research.
2. General criteria for research evaluation.
3. Elements of the study of the change process.
4. Methods and techniques for studying the change process components.
5. Criteria of adequacy for these techniques.

---

[10] G. B. Johnson, "A Method for Evaluating Research Articles in Education." Journal of Educational Research 51:149-51; October, 1957

[11] American Institute of Research, "A Procedure for Evaluating Graduate Research on the Basis of the Thesis." Pittsburg: October 1955.

[12] W. J. Gephart, Development of an Instrument for Evaluating Reports of Educational Research. Washington, D.C.: U.S.O.E. Cooperative Research Project Number S-014, 1964.

A PLAUSIBLE LOGIC FRAMEWORK FOR EDUCATIONAL RESEARCH

The literature on the research process presents considerable agreement regarding the components of the research process. The student of the research process has little difficulty identifying components of (1) problem identification and development, (2) evolution of hypotheses, (3) evaluation and synthesis of previous research, (4) designing the specific study, (5) analyzing the data accumulated, and (6) derivation of the conclusions and implications. Although understanding of each of these is important to the conduct of research and to its evaluation, the discussion which follows focuses on the premise that each research effort is in itself a logical argument.

As knowledge builds up about a substantive area, each piece of research attempts to provide some direction for further expansion of knowledge. That is, when an unknown or a problem is encountered, several possible solutions are identified, each of which presents an hypothesis to explain the unknown or solve the problem. The tests of these hypotheses assist in the most efficient movement to the next "fork in the tree."[13]

The hypothesis in a given study then is "a conjectural statement about the relationship between two or more variables."[14] The

---

[13] J. R. Platt, op. cit.

[14] F.N. Kerlinger, Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1964.

focus of each study is the establishment of the truth of the hypothesis. Does empirical evidence support the validity of a theoretically evolved hypothesis?

This reasoning form differs from formal logic where observations regarding the truth of an antecedent are used to infer truth of a consequent. For example:

Major Premise:	If I live in Oconomowoc (antecedent); then I live in Wisconsin (consequence).
Minor Premise:	(a) I live in Oconomowoc, or (b) I do not live in Oconomowoc.
Conclusion:	(a) The consequence is true, I live in Wisconsin, or, (b) No conclusion regarding the consequences.

If forced to observe on the consequence, the possibility of a positive conclusion is removed.

Minor Premise:	(a) I live in Wisconsin, or (b) I do not live in Wisconsin.
Conclusion:	(a) No conclusion regarding the antecedent, or, (b) The antecedent is false, I do not live in Oconomowoc.

Hypotheses in the social sciences are generally not directly observable. Thus, the researcher is compelled to consider the hypothesis as the antecedent in a syllogistic major premise which, if a true statement, would result in certain observable consequences. The form of logical inference suggested by the mathematician,

Polya,[15] is appropriate to infer an answer to the question, "Is the hypothesis true?"

Major Premise:	If A (hypothesis) then B (consequence) is a true statement.
Minor Premise:	B (the consequence) is observed.
Conclusion:	The truth of A is supported.

In proposing the application of this "plausible inference pattern" to educational research, Raths indicates the need for the insertion of a qualification prior to the minor premise.[16] This qualification is necessitated due to awareness that variables other than those specified in the hypothesis may affect the degree to which the consequences are observed. For example, performance at a learning task may be due to: age, sex, prior knowledge, attitudes, etc., as well as--or even rather than--being due to a specific treatment. In the research process this qualification is evident in the form of control of such extraneous and error variances. Thus, the inference pattern skeleton of the research process is outlined as follows:

Major Premise:	If A (hypothesis) then B (consequence) is a true statement.
Qualification:	B occuring without A being true is hardly credible due to controls employed. (Amount of control can be equated to the number of alternative hypotheses eliminated.)

---

[15] George Polya, presented in an unpublished lecture, University of Wisconsin-Milwaukee, December, 1963.

[16] J. Raths, Unpublished paper presented at the American Educational Research Assocication Annual Conference, February, 1964.

Minor Premise: B (consequence) is observed.

Conclusion: The truth of A is strongly supported. (The strength of the support is proportional to the amount of control.)

The components of this inference pattern relate to the components of the research process. The researcher through the identification, definition, and delimitation of a problem, develops a theory consisting of what is known and what is suspected. The latter need to be tested in the form of hypotheses. The determination of the data necessary to test the hypothesis is arrived at by deducing the consequences which may be observed if the hypothesis is true. Thus, through the problem, hypothesis, related research components of the research process, the major premise is established. The design and data analysis components of the research process equate to the qualification, in that design generally refers to the plans made to ensure the collection of the most relevant data on the consequence as a test of the hypothesis and data analysis techniques are also employed as controls. The findings resulting from the analysis of the data are the specifics of the minor premise, and finally, the conclusion component of the research process coincides with the conclusion in the plausible inference pattern.

GENERAL CRITERIA FOR RESEARCH EVALUATION

Criteria for evaluating research have been stated by numerous individuals. These may be listed in two categories: lacking or having data available regarding validity and/or reliability.

Examples of checklists lacking such data are those proposed by Symonds,[17] Van Dalen,[18] Farquahar and Krumboltz,[19] and Mouly.[20] Those proposed by the American Institute of Research,[21] Johnson,[22] Wandt,[23] and Gephart[24] have endured some empirical assessment. However, similar to any group of psychological measuring devices, they seem to differ in quality for the purpose at hand. With the exception of Wandt's work which is yet to be reported, the instruments in this latter group are critiqued below.

Johnson's instrument is exceedingly brief. It consists of eleven items: two each for evaluating the problem, materials, and subjects; three for the method of procedure; and one each for evaluating results and conclusions. The nature of the items further reduces the value of Johnson's instrument. For example, about the problem the instrument asks, "Is it clear? 1-2-3-4-5."

---

[17] P. Symonds, "A Research Checklist in Educational Psychology." Journal of Educational Psychology 47:100-9; February 1956.

[18] D. B. Van Dalen, "Research Checklist in Education." Educational Administration and Supervision. 44:174-81; May 1958.

[19] W. W. Farquahar and J. D. Krumboltz, "A Checklist for Evaluating Experimental Research in Psychology and Education." Journal of Educational Research 52:534-5; September 1959.

[20] G. J. Mouly, The Science of Educational Research. New York: American Book Company, 1963. p. 503-4.

[21] American Institute of Research, op. cit.

[22] G. B. Johnson, Jr. op. cit.

[23] E. Wandt is chairman of an ad hoc committee on research evaluation for the American Educational Research Association.

[24] W. J. Gephart, op. cit.

The focus of a judgment on a term such as "clear" is asking for the application of a rubber yardstick. What is clear to one person may be obtuse to another. The other items focus on terms of the same nature: "significance," "authoritative sources," "large enough samples," "adequate," "orderly and systematic," "proper and modern techniques," etc.

Johnson was able to obtain significant agreement when fourteen of his students (late in a course on educational research) and four of his colleagues used his evaluative instrument. Private correspondence with Johnson indicates that the course was devoted to the definition of the terms cited above. Thus, it would seem that when commonality of definition of the research process exists, agreement among evaluators can be obtained in assessing research adequacy. It is doubtful though that this brief instrument can provide the necessary definition.

The project reported by the American Institute of Research (AIR) substantiates the conclusion drawn from Johnson's work.[25] Although the instrument in this study is considerably longer, it still uses terminology which needs definition. The AIR instrument is a distinct improvement in the responses requested of the rater in that it lists actions to be completed in the research process and requests two responses: did the action occur; and, does that occurrence contribute or detract from the value of the research.

---

[25] G. B. Johnson, op. cit.

Gephart,[26] following some work with Clark, Guba, and Smith,[27] also focused upon actions inherent in the research process. Numerous texts were analyzed to identify actions to be taken in the research process. An attempt was made to avoid the undefinable terms such as were found in the above mentioned studies. The response pattern of the AIR instrument was employed. That is, occurrence and value ratings were requested for each item. A Cooperative Research Program Grant made possible the use of ten competent judges in the establishment of (1) the applicability of each item, (2) the comprehensiveness of the instrument, and (3) the interrater reliability for evaluations of reports of research in professional journals. Significant agreement was found both within and among the five jurors who were research design and methodology experts and the five substantive experts.

These efforts seem to imply at least that when judges of research adequacy employ the same set of definitions of the research process, there is reliability. They can agree on the adequacy of a specific research. It has been suggested that, when jurors employing these instruments have sufficient commonality of experience and training, the instrument serves as a reminder function, calling to the evaluator's mind all of the factors to be

---

[26] W. J. Gephart, op. cit.

[27] David L. Clark, Egon G. Guba, and Gerald R. Smith, "Functions and Definitions of Functions of a Research Proposal or Research Report." Unpublished mimeo, Columbus, Ohio, The Ohio State University, 1962.

considered. If such is the case, assessment of research adequacy could proceed using Symond's or Van Dalen's checklists or the instruments developed by the AIR or Gephart, for all of them attempt comprehensive coverage of the research process.

Despite differences in terminology, type of response and specific focus of the individual items in the above checklists, they have in common the research process components found in every text on the research process. All indicate that the evaluation task should focus on (1) the problem studied, (2) the hypotheses tested or questions asked, (3) the related literature surveyed, (4) the design of the study, (5) the analysis of the data, and (6) the conclusions and implications drawn from the study. Thus, it is proposed that the assessment of the adequacy of research on change in education should be based on (1) the general criteria of research adequacy, and (2) the criteria relevant to research activities specific to techniques of research on change. The enumeration of general criteria is presented next. Criteria having relevance only to research on change will be presented after brief statements on the elements, methods, and techniques for such study.

Evaluative Criteria for the Problem Component

In discussing the problem component, the literature on the research process evidences some degree of agreement on the following activities:

1. The establishment of the existence of a problem.[28, 29]
2. The identification of the factors or variables inherent in the problem.[30, 31]
3. The relating of the problem to its antecedents.[32, 33]
4. The identification of the limits in the study of the problem.[34, 35]

---

[28] David L. Clark, Egon G. Guba, and Gerald R. Smith, op. cit.

[29] G. D. McGrath, James J. Jelinek, and Raymond E. Wochner, Educational Research Methods. New York: The Ronald Press Company, 1963. p. 24

[30] George J. Mouly, op. cit.

[31] Debold B. Van Dalen, Understanding Educational Research. New York: McGraw-Hill Book Company, 1962. p. 23.

[32] G. D. McGrath, James J. Jelinek, and Raymond E. Wochner, op. cit. p. 52.

[33] David L. Clark, Egon G. Guba, and Gerald R. Smith, op. cit.

[34] G. D. McGrath, J. J. Jelinek, and R. E. Wochner, op. cit. p. 52.

[35] D. B. Van Dalen, op. cit. p. 52.

5. The description of the significance of the research.[36, 37, 38, 39]

6. The description of the goals the investigator intended to achieve.[40]

7. The definition of terminology utilized in the study.[41, 42]

It is proposed that these seven points can be synthesized into four criteria for evaluating the problem.

1. Does the researcher establish the existence of a problem?

2. Does the researcher develop a theory or conceptual framework for the problem?

3. Does the researcher describe the specific goals to be achieved?

4. Does the researcher state the limits within which the study is conducted?

---

[36] Walter R. Borg, Educational Research: An Introduction. New York: David McKay Company, Inc., 1963. p. 31.

[37] D. B. Van Dalen, op. cit. p. 125.

[38] T. A. Lamke, "Primer in Research: Lesson I. Identifying and Defining the Problem," Phi Delta Kappan, 38:127-33; January 1957.

[39] D. L. Clark, E. G. Guba, and G. R. Smith, op. cit.

[40] Ibid.

[41] Carter V. Good, Introduction to Educational Research: Second Edition. New York: Appleton-Century-Crofts, Inc., 1963.

[42] John W. Best, Research in Education. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1942.

Several types of situations are described which should be helpful in deciding the establishment of the existence of a problem. Van Dalen[43] and McGrath, Jelinek, and Wochner[44] describe problems as either (1) the adaption of a means to an end, (2) the lack of understanding of the character of an object or event, or (3) the existence of an unexpected event. Clark, Guba, and Smith[45] use different terminology and add a category as they indicate that a researchable problem is an anomaly, an uncharted area, an unverified "fact," or the existence of conflicting evidence. It is here proposed that the establishment of a problem has been accomplished if the researcher documents the existence of one of these situations.

Throughout the research process literature there is a concern expressed for the lack of an integral role for theory. The importance of theory has been most eloquently stated by Platt[46] as he indicates that a major difference between those sciences that make rapid strides and those that languish in their data is the use of "strong inference." This he indicates evolves from the theoretical construction of a logical inference tree,

---

[43] D. B. Van Dalen, op. cit.

[44] G. D. McGrath, J. J. Jelinek, and R. E. Wochner, op. cit.

[45] D. L. Clark, E. G. Guba, and G. R. Smith, op. cit.

[46] J. R. Platt, op. cit.

a tree where the forks of the branches represent alternative solutions to problems impeding man's progress or knowledge. The systematic testing of these alternatives adds to our understanding. The failure to use some guiding conceptual framework leaves us collecting data which has unknown relevance.

The appropriate question to facilitate research evaluation is how does one build the conceptual framework or theory so desired in research? It is here believed that this is initiated through the activities of identification of (1) the variables known and/or suspected to be operant in the problem area, (2) the relationships among these variables again including both the known and suspected, and (3) the educational, social, and scientific antecedents of the problem situation. The theoretical base or conceptual framework is completed when the researcher is able to structure and state a set of assumptions which will enable him to conjecture as to what is and where in the scheme of this is the problem.

It should be pointed out that in the last sentence the word "assumptions," plural, was used. These assumptions are the focus for research, for the advancement of our science requires the movement of a point from the category of assumption to the category of fact. That is, our end is being able to know what variables are involved rather than accepting their possible involvement--knowing the relationships of these variables rather

than accepting the possiblity and nature of their relationship. The acceptance of this point establishes the need for the specification of the objective to be achieved in a given study for two objectives, to identify variables and to test relationships are implied each of which may be subdivided into an array of goals.

The final criterion, the statement of the limits within which the study is conducted, relates both to the theory woven about the problem and to the setting in which the problem is studied. As implied in the above discussion, the hypothesis to be tested in a study is derived from one or more of the assumptions in the theory. The existence of other assumptions sets limiting conditions on the study which must be considered as the research progresses. It is also possible that the site of the test of the hypothesis, both in terms of time and physical characteristics, provides some limits to the absolute solution of the problem. Thus, their identification is mandatory as an aid to the reader's interpretation of the study.

In closing this discussion on the evaluation of the problem, attention is called to the absence of a criteria of significance or justification for the research. This omission is by design for it is here believed that the establishment of the existence of a problem and the structuring of a theoretical framework descriptive of it provide sufficient justification for its study.

Evaluative Criteria for the Hypothesis Component

Writings on the role of hypothesis in research project an almost human capacity to something that is little more than a collection of words. For example, one can find statements that hypotheses

1. . . . provide direction to research.[47]
2. . . . prevent the review of irrelevant literature or the collection of useless data.[48]
3. . . . sensitize the investigator to certain aspects of the stiuation which are relevant.[49]
4. . . . is required to provide a framework for stating the conclusions in a meaningful manner.[50]
5. . . . serves as an intellectual lever by which investigators can pry loose more facts to be fitted into other more conclusive explanations.[51]

Skipping through these statements conjures up the picture of a little genie that appears magically and whispers in the investigator's ear, "Don't read that study. It's irrelevant."; Who grabs the investigator's pencil and shouts, "Don't record these data. They're useless!" Then, by magic, the hypothesis genie

---

[47] G. J. Mouly, op. cit. p. 89-90.

[48] Ibid.

[49] Ibid.

[50] D. B. Van Dalen, op. cit. p. 156.

[51] Ibid.

alters his form to become a solid steel skeleton upon which flashing neon conclusions are fastened. Again an alteration, and our faithful hypothesis is the longest and strongest of crowbars. Would that researchers could find such a dandy companion.

Enough of the dreaming; if one is to evaluate hypotheses, it is imperative that he know what they are as well as what they are not. Definitions of the term range in sophistication from Hillway's statement that a hypothesis is ". . . a reasonable guess or supposition based upon the evidence available at the time the guess is made,"[52] to Guba's statement that

> Within the framework of a theory, hypotheses are deductions following from and logically consistent with the assumptions on which the theory is based.[53]

The statistician provides a different focus in stating, "Hypotheses, whether statistical or research, are usually concerned either with differences or deviations."[54] This writer prefers Kerlinger's attempt at synthesizing all of the above as he defines a hypothesis as ". . . a conjectural statement of the relation between two or more variables."[55]

---

[52]Tyrus Hillway, Introduction to Research: Second Edition. Boston: Houghton Mifflin Company, 1964. p. 123.

[53]Egon G. Guba, "The Writing of Proposals," in Research in Educational Administration, edited by Stephen P. Hencley.

[54]Quinn McNemar, Psychological Statistics. New York: John Wiley and Sons, Inc., 1955. p. 61.

[55]F. N. Kerlinger, op. cit. p. 20.

With this definition of a hypothesis let us return to the magical claims in the literature. It should be patently clear that a statement is unable to direct, prevent, sensitize, etc. It should also be clear that these are necessary aspects of research. That is, the research must have direction. The researcher must classify and categorize the irrelevance of data. He must establish a framework for conclusions. The hypothesis stated in a research is only the mode used by the human to state those aspects he has worked through.

Accepting this argument, four criteria seem relevant to the task of evaluating a hypothesis.

1. Does the hypothesis state or directly imply the existence of two variables?[56]
2. Does the hypothesis state or directly imply a relation between the variables?[57]
3. Are the variables empirically observable?[58]
4. Is the hypothesis based in a theory or a body of previously established knowledge?[59]

Point 4 in the above list bears some elaboration. One contributor to the snail-like rate of progress due to educational

---

[56] Ibid.

[57] Ibid.

[58] G. J. Mouly, op. cit. p. 92.

[59] Ibid. p. 91.

research is the tendency to empirically approach a single hypothesis. In contrast the physical scientist typically starts out with sets of hypotheses which he systematically works his way through. He seems to ask himself, "What could have caused this?" and answer hypothetically, "It could have been A, or B, or C . . . ." The test of a single one of these hypotheses invites little or no advancement through a no-significant-difference finding or the failure to identify multiple causation. Thus, the failure to base the evolution of a research hypothesis either in theory or substantial body of knowledge from which rival hypotheses can be or are evolved reduces the effectiveness of a given study.

This set of criteria rejects one item frequently found in the literature. Mouly enunciates this one clearly as he states, "A good hypothesis must be stated as clearly and concisely as the complexity of the concepts involved will allow."[60] This seems to structure the evaluation of research on the matter of literary style rather than on actions in the research process.

The potential research evaluator ought to ask at this point what about the case in which a hypothesis is not explicitly or implicitly stated. Do we reject as research the situation in which a concern focuses our attention on an area in which the quantity and quality of existing knowledge precludes

---

[60] Ibid.

hypothesizing? The "uncharted area" problem category is a case in point. If one considers a field of study in which no prior work has been done, it is unlikely that he is able to specify the constructs that may hypothetically explain the phenomena. In this situation he needs information which would describe the number, nature, and relationship of these concepts. In other words, he wants to know what are the variables that are involved, what is their nature, and what are the relationships between variables. In this respect questions can be employed to give "direction to a study," "prevent the collection of irrelevant data," "provide a framework for conclusions," and so on through all the claims made for hypotheses. Thus, the criteria for evaluating questions should include

1. Does the question seek either the identification or nature of variables in a given problem?
2. Is the variable in each question observable?
3. Is the question related to the existing body of knowledge?

The presentation of these criteria for evaluating questions in research speaks directly to Platt's[61] concern for asking the "crucial question" by stating that a question should either seek the identification of variables or their description. It further proposes that if enough is known about a problem that a researcher can conjecture on the existence and relation of variables, a hypothesis is warranted.

---

[61] J. R. Platt, op. cit.

Evaluative Criteria for the Review of Related Literature[62]

In the discussion of both the problem and the hypothesis or question components above, definite implications for the review of literature have been enunciated. This segment will attempt their explication.

The individual who reads quantities of research reports frequently is in agreement with Lindvall's[63] judgment that all too often the review is not an integral part of the study. It is here proposed that this difficulty is a direct result of the general trend Lake[64] finds, i.e., that the majority of researchers are raw empiricists in contrast to hypothetic-deductivists. To the latter, knowledge is cumulative. Thus, the use of what is known to set the theoretical framework gives meaning and relevance for a related literature review in a report. Failure to see such a purpose makes the review almost an academic task of producing a lengthy annotated bibliography and/or proving the uniqueness of his study. Both are rejected by Lindvall[65] as central to the study.

---

[62]Much of this discussion is adapted from the Clark, Guba, and Smith outline; op. cit.

[63]C. M. Lindvall, "Review of Related Research." Phi Delta Kappan 40:179-80; January 1959.

[64]K. E. Lake, op. cit.

[65]C. M. Lindvall, op. cit. p. 179.

The purpose of reviewing the literature then is for the development of a theoretical or knowledge base upon which the substance and methodology of a given study are built. Criteria for evaluating this achievement have already been expressed under headings above. There are, however, specific activities involved that should be elaborated here as the basis for establishing specific criteria.

If the researcher is to facilitate growth on the part of any audience, he can do so by helping them to know the setting. Thus, by listing the extent of the review and the specific bibliographic references found relevant, the researcher contributes to the definition of the setting in which the problem is studied.

To accept as complete, stopping with mere listing of related literature or providing what Monroe and Englehart call a "classified annotated bibliography,"[66] is to declare this entire paper and much of the focus of this conference as unnecessary. The differential value of various researches has been substantially documented. Thus, if a review is to be of value in building a theory of knowledge base, the strengths and weaknesses of each article included must be identified.

---

[66] W. S. Monroe and M. D. Engelhart, The Scientific Study of Educational Problems. New York: The Macmillan Company, 1936. p. 437.

The final activity for evaluating the review may relate to the manner in which the problem statement is presented and the design is structured just as well or better than under a separate rubric, related literature. This activity is the synthesis of what is and is not known about the subject.

Each of the items for evaluating the review also enunciate the review's topical emphasis. Here two areas are proposed: the substantive area and the methodological area. Through the review the individual should be presented with a synthesis of what is and is not known about the subject at hand and about the proposed method for studying that subject.[67]

The specific criteria could be stated as follows:

1. Does the research report present a list of the studies completed in both the substantive and methodological aspects of the problem?

2. Does the research report present a critique of the studies listed?

3. Does the research report include a synthesis of what is known in both the substantive and methodological aspects of the problem?

Evaluative Criteria for the Design

The discussion under this rubric will be restricted to the definition of the term and brief recognition of some criteria

---

[67] D. L. Clark, E. G. Guba, and G. R. Smith, op. cit.

applicable to all methodologies. This discussion will be augmented in a later section devoted to criteria for evaluating research techniques specific to the study of change.

Neither those engaged in research nor those observing their work attempt to refute Barr, Davis, and Johnson's statement that

> Educational research is a complex activity; only through the most meticulous specifications can the many factors that need to be kept in mind be controlled at the proper time.[68]

Thus is the justification of the design component of the research process.

Design in this context is defined as that planning in which the researcher engages to insure the accumulation of the most powerful conclusions about the nature of the problem. Journals have long conveyed the assumption that a consumer should be informed of these plans through the inclusion of a procedures or design section in their format.

Lindquist presents a cogent statement in describing the ingredients of an experimental design.

> The important decisions to be made in planning the experiment are concerned with: (1) the definition of the "treatments," (2) the selection and exact definition of the population to be investigated, (3) the selection of

---

[68] A. S. Barr, R. A. Davis, and P. O. Johnson, Educational Research and Appraisal. Chicago: J. E. Lippincott Company, 1953. p. 309.

> a criterion, (4) the identification of the factors to be controlled, (5) the final restatement of the problem, and (6) the selection of a specific experimental design.[69]

If broadly interpreted, many of these are applicable to descriptive and historical methodologies. For example, items (1) and (2) in this list are important to both historical and descriptive research. The historian is interested in determining what is the pattern of events (or treatments) and the strength of their contribution to an historical event. This is documented by Travers as he states, "Historical studies usually begin with a delimitation of the general category of events that is to be reconstructed."[70] The descriptive researcher is interested in the status of a particular group at a given time. The reason for his interest, that is the pattern of events or circumstances which have made this group the focus of his interest equates to the experimental term "treatment."

Another example of the extension of Lindquist's ingredients of design is in the area of criterion selection. In any research, descriptive, historical, or experimental, the decision

---

[69] E. F. Lindquist, Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Company, 1953, p. 7.

[70] R. M. W. Travers, An Introduction to Educational Research: Second Edition. New York: The Macmillan Company, 1964, p. 115.

must be made regarding what is acceptable evidence either for testing the hypothesis or answering the questions relevant to the problem.

The specific decision points upon which research planning focuses are the population, the sample, the variables involved, the controls necessary, the data collection techniques, and the analysis procedures Previous sections have discussed the justification of the focus on variables and their relationship. If through research we attempt to make statements with any generality, our plans must include a careful focus on population. What are the characteristics of the units in the population(s)? (The parenthetical plural is extremely important in some studies; e.g., McNeil[71] forgets that he has a population of students and a population of teachers.) Without careful definition here the researcher leaves no avenue for applicability.

Sampling in a given population not only enables the researcher to focus upon a group in which he can efficiently conduct his study but also has relevance to the analytical model. Certain sample characteristics support the Neyman-Pearson model, others upport a Bayesian approach. It is of little value here to debate the adequacy of these theories as such requires far more accomplished statisticians than I. However, it is important that we focus on the need

---

[71] J. D.McNeil, "Programmed Instruction Versus Usual Classroom Procedures in Teaching Boys to Read," American Educational Research Journal 1:113-121; March, 1964.

for information about the sample(s) utilized in order that the design might be evaluated.

The importance of the data collection technique(s) in a study are generally accepted. If a researcher is to make an inference regarding the truth of his hypothesis, he must collect observational data on suspected consequences. The objectives in any data collection technique are ". . . to provide accurate observation, to eliminate observer bias, and to extend and quantify the observations of the human researcher."[72] Three concerns seem imperative. Are the techniques valid for measuring the consequences predicted? Is there consistency in the measurement? And finally, are the techniques objective? The general acceptance of these points is so great that it was a source of amazement to find that criteria of instrument reliability and validity did not discriminate between good and poor research reports in Gephart's study of a research evaluation instrument.[73]

Although procedures for analysis of the data are decisions inherent in the design, their importance as an aspect of research warrants their treatment as a major component of the research process rather than being subsumed under the heading of design.

---

[72] Arthur J. Bachrach, Psychological Research: An Introduction. New York: Random House, Inc., 1962, p. 32.

[73] W. J. Gephart, op. cit.

Given the above, general evaluation of the design rests on the adequacy which can be attributed to the anwers to the following questions.

1. Does the research report define the population of people, things, or occurrences inherent in the problem?

2. Does the research report describe the sample selection procedures and/or the characteristics of the sample?

3. Does the research report operationally define the variables studied and the variables known to be associated in the problem?

4. Does the research report describe the controls employed to counter the effects of the latter group of variables?

5. Does the research report specify optimally valid and reliable data collection devices or techniques?

Evaluative Criteria for the Analysis of the Data

Systematic analysis of the accumulated data is imperative in order to determine inherent facts and meanings not necessarily apparent in casual examination. A negative example makes the point. Hart[74] asked counselor educators and teachers to rank the importance of forty-one tasks performed by elementary school counselors. His observation that the relative importance attached to several of the tasks by the two groups differed greatly led

---

[74]R. N. Hart, "Are Elementary Counselors Doing the Job?" The School Counselor 9:70-2; December 1961.

him to conclude that counselors could not make both groups happy. Had he calculated a correlation between the two ranks he would have found agreement (Rho = .796) significant at the .001 level. Rather than his pessimistic conclusion, he should have acknowledged the significant agreement among the two groups.

Stanley's[75] discussion of research reports in Volume I of the American Educational Research Journal identifies additional cases of analytical inadequacies.

Best has stated the outcomes of the data analysis.

> . . . the research process is not complete until the data are organized and analyzed, and significant conclusions are derived. These conclusions will be based upon comparisons, contrasts, or relationships of one kind or another.[76]

Thus it would appear that in the analysis of the data a statistical description of the data and the statistical significance of these data are vital.

Much has been written which specifies which statistical procedure should be employed on a given set of data. Three notable attempts at synthesizing this literature provide assistance in determining which statistic is appropriate. These can

---

[75] J. C. Stanley, "The Improvement of Educational Experimentation." Mimeographed paper read to Seventh Annual Phi Delta Kappa Symposium, Madison, Wisconsin, 1965.

[76] J. W. Best, op. cit. p. 103.

be found in the tables constructed by Senders,[77] Siegel,[78] and Tatsuoka and Tiedeman.[79] In each case a grid has been constructed through which the appropriate statistic can be identified by determining the number of variables involved in the analysis, their scalar nature, and the relationship between the samples (dependent or independent samples).

Given the above discussion, evaluation of the data analysis in a given study rests on the nature of the answers to the following questions.

1. Does the research report systematically organize the accumulated data?

2. Does the research employ appropriate statistical procedures in analyzing the data? (Appropriate herein is defined by the scalar nature of the data and the design employed.)

Evaluative Criteria for Conclusions

Many have expressed concern for what goes into the conclusion component of the research process. This concern ranges from the contents of the conclusion statement to its form. The

---

[77] V. L. Senders, Measurement and Statistics. New York: Oxford University Press, 1958.

[78] S. Siegel, Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, Inc., 1956.

[79] M. M. Tatsuoka, D. V. Tiedeman, "Statistics as an Agent of the Scientific Method in Research on Teaching," in Handbook of Research on Teaching. N. L. Gage, Editor, Chicago: Rand McNally and Company, 1963.

first is exemplified by Hillway's statement regarding the kinds of statements that might be admissible.

> These are: (1) a basic assumption, (2) a statement of fact, (3) the writer's opinion, and (4) the opinion of an authority in the field.[80]

The evaluation of conclusions would be facilitated if each of these types of statements is clearly identified. Van Dalen's words attach the evaluation of conclusions to the plausible logic inference pattern described earlier. He indicates that a conclusion cannot be drawn regarding the truth of the hypothesis until

> . . . it meets all of the following requirements: (1) all the factual evidence collected in the empirical tests corresponds with the consequences (of the hypothesis); (2) the test situation adequately represents the essential factors expressed in the consequences; and (3) the consequences are logically implied in the hypothesis.[81]

Synthesizing these three statements defines a conclusion as a statement about the truth value of the hypothesis given the conditions of the specific study.

If a research is to contribute to making cumulative the body of knowledge or to the evolution of theory, the research should present a statement of implications. That is, if in the analysis of a

---

[80] T. Hillway, op. cit. p. 137.

[81] D. B. Van Dalen, op. cit. p. 139.

specified set of data, support is garnered or lost for a given hypothesis, then a discussion which speaks to the meaning of this conclusion for our professional growth is imperative. Does this specific study strengthen certain theoretical assumptions, or does it imply modification in the theoretical base and suggest needed research?

Given acceptance of the above statements, the following questions are posed as criteria for evaluating the conclusion component of a research report:

1. Does the report state whether the findings firm or disconfirm the hypotheses?
2. Does the report state the conclusions drawn from the findings?
3. Are the conclusions drawn from without going beyond the data?
4. Does the report describe implied modification in theory raised by the conclusions?
5. Does the report state specific problems raised by the investigation that require additional research?

ELEMENTS OF THE STUDY OF THE EDUCATIONAL CHANGE PROCESS

To set the stage for a discussion of criteria specific to the assessment of methodological adequacy it is believed necessary to attempt a definition of the field for such research. The

synthesis of statements by Rogers[82] and Walton[83] provides the basis for this definition task.

Rogers' work highlights three variables inherent in the study of change, the innovation, the target unit, or that which is to be changed, and the initiating unit or change agent. In the discussion which follows these variables will be included under the rubric "actor variables."

As one examines research on change it becomes obvious that there are interactions between these actor variables. That is, an innovation with a certain set of characteristics is more acceptable or less acceptable to target units of different character. A change agent of one type may be effective with one innovation and not with another, or with one target unit and not another. In statistical language, the field of study focuses on the description of, and the assessment of, the main and interaction effects of the three actor variables.

A second set of variables, "action variables," seems implied as Rogers also describes an adoption process or an action sequence. His presentation seems limited, however, as it is designed from the target unit vantage point. As such, it implies but also tends

---

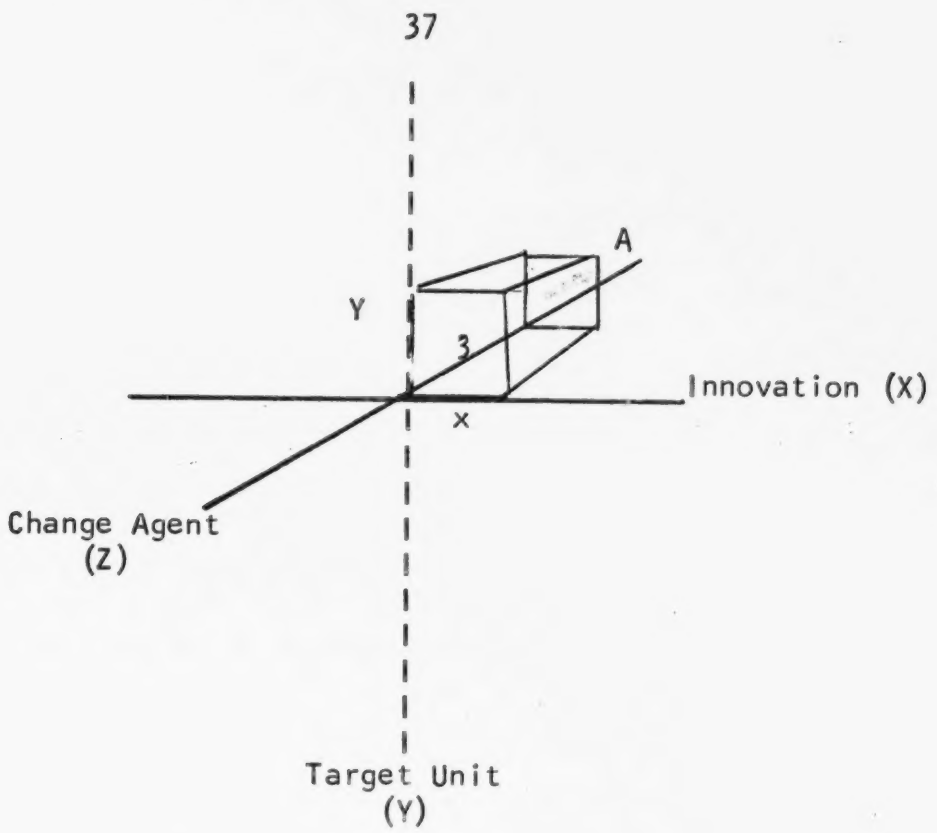
[82] E. Rogers, Diffusion of Innovations. New York: The Free Press of Glencoe, 1962.

[83] R. E. Walton, "Two Strategies of Social Change and Their Dilemmas." The Journal of Applied Behavioral Science. 1:167-79; April-May-June 1965.

to obscure the concept of action on the part of the change agent. Walton's discourse highlights the interaction between two change agent action variables. These are actions encompassed in either a power strategy or an attitude strategy and his discussion of them is thought provoking. The purpose here is not to question these strategies as the only possible actions but to highlight their existence and interaction. The description of these action variables and the assessment of their effect of the change process, either individually or in concert, provides a second focus in the definition of the field of study.

Just as there is suspected interaction between actor variables and between action variables, so too there should be suspected interaction between actor and action variables. That is, change in a given target unit may best be accomplished through a given action. Or, change may be facilitated if a specific behavior is employed in a situation involving an innovation-target unit interaction.

Given the above points a four dimensional model can be constructed. As four dimensionality is impossible to picture graphically, the following build up of the model is presented. It is possible to conceive of the three actor variables as each contributing a main effect to the process of change in an educational institution and their contribution through two or three way interactions. Thus three axes in the model are set by the actor variables.



Any given innovation may contribute x units to the amount of change that occurs in the system. A specific target unit, a school, an educational level or discipline, a specific teacher contributes y units. (Here may be the propitious point to insert the possibility of negative change.) Finally, the change agent adds (or detracts) z units of change due to his nature, position, relation to the target unit, etc. Point A then represents the amount of change expected in a system from the static existence of an innovation, a target unit, and a change agent.

When the action variable--that is, a power strategy of legisation, remuneration, etc., or an attitude strategy of interpersonal involvement, education, etc.,--is inserted into the study of the change process, a fourth dimension is needed. In other words when an innovation, provision of counseling for students, is offered to a target

unit, the secondary school, by a change agent, a professional organization, change of differing quality and quantity will result from differing actions, legislative lobbying versus educating teachers as to the need for such a service. The inclusion of all four factors and their interrelationships can be affected through a four-dimensional model in which the actor variables account for three of the dimensions and the action variables, the other. Such a model focuses our attention on the following:

1. The actor variables: What are they and what variance exists on each?

2. The action variables: What actions or activities are involved and again what is the variation possible on each?

3. The interrelationships between actor, action, and actor-action variables: What interrelationships exist? What are the effects of one upon another?

The researcher can interpret this model in a manner which suggests research activities. First, it seems imperative that the variables be identified and that their characteristics be understood. If we are to build a model of change, we must know its constituents. Second, if our model is to become a theory, we must know these constituents well enough to at least conjecture about their relationships. Thus it would seem that historic and descriptive studies are important in setting the model and that experimental studies are valuable in testing the relationships that exist.

METHODS AND TECHNIQUES FOR STUDYING THE CHANGE PROCESS COMPONENTS

The terms "research methods" and "research techniques" are too often used interchangeably in the literature. In the discussion which follows, "method" refers to the investigatory strategies: historical, descriptive, or experimental study; while "techniques" refer to the specific actions taken in a given study. This latter area includes techniques of sample selection, treatment, data collection, and data analysis.

The description of the three research methods listed above typically places them on a continuum of time. For example, Best states,

> Historical research describes what was.
> Descriptive research describes what is.
> Experimental research describes what will be when certain factors are controlled.[84]

A second dimension for categorizing research methods is presented by Best in the same discussion. In this he discusses the techniques typically employed in each method. For example, the historian attempts to identify "primary, original, or first-hand sources of information,"[85] for the purpose of understanding change. The descriptive researcher engages in the accumulation and analysis of data for the purpose of describing status.

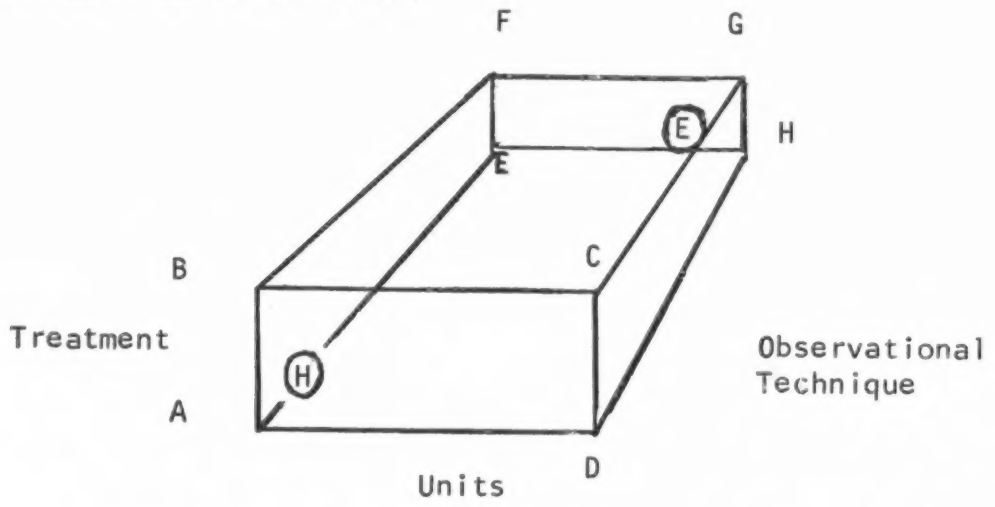
---

[84] J. W. Best, op. cit. p. 12.

[85] Ibid.

The merger of the above two dimensions for defining research methods induces confusion. If one is interested in the extent to which a specific innovation--let us say providing counseling opportunities in high schools--had been adopted in 1950, he is attempting to go back into the past to determine what was. Further, he must seek out first hand sources of information to conduct a valid study. At the same time, he is accumulating and analyzing data to describe status. Is he engaged in a historical or descriptive study? An ex post facto experimental design presents some of the same conflict. Case studies also add to the confusion.

It is here proposed that, rather than a conceptualization of research methods as categories, greater clarity may be obtained through analysis of methods according to the amount of control an investigator has in generation of data. Three factors, the units or subjects involved in the population and sample, the treatment(s) of these units, and the data accumulation techniques employed, structure the generation of data. Thus research method exists in a cubic model as shown below.



In this conceptual scheme historical research exists at the lower left hand corner (A) of the three axes. Here data was generated through unspecified observational techniques on an unselected set of units which experienced some uncontrolled treatment(s). The true experiment involves the selection of subjects on a random basis, careful control over the treatment, and the selection or construction of valid and reliable observational techniques. Thus the true experiment exists at the opposite corner of the cube (G). Quasi-experimental studies as described by Campbell and Stanley[96] exist when the investigator has control over the treatments and the observational techniques employed but lacks control over the units involved. As such quasi-experimental studies are located on the ABFE face of the cube. The "better" the quasi-experiment, i.e., the greater the control over treatment and the more valid and reliable the measuring technique, the closer the study is located to point F.

The descriptive research method omits control over treatment. In such a study, the researcher has the power to structure the inclusion of units and to utilize valid and reliable observational techniques. Thus the descriptive study locates somewhere on the AEHD face of the cube. Again the "better" the study, the closer it may be located to a corner of the cube, in this case H.

---

[86]D. T. Cambell and J. C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching" in Handbook of Research on Teaching. N. L. Gage, Editor; Chicago: Rand McNally, 1963.

Conceptualization of research methods through this cubic model facilitates the understanding of the contrubution of each method. Perhaps, in light of this, we might avoid the frequent valuing and devaluating of research by methodological types and base our judgments on the adequacy of the specific techniques employed in a given study.

The application of each of these methodologies can contribute significantly to knowledge of the process of change in education. For example, information about the process of change could be gained through historical study of the spread of an innovation; descriptive studies on the nature of the innovative activities currently found in educational institutions also are valuable. In essence this is the methodology employed by Mort and his associates. Ascertaining the effect of one variable on another, the experimental method provides a significant means for developing understanding and making predictions regarding the action-actor variable interactions. Thus, the adequacy of the several research methods for the study of the process of change depends on the objectives of the investigation.

The term "research technique" here means those activities of subject selection, treatment administration, and data collection, evaluation and analysis employed in a given study regardless of general research method employed. An examination of the research reports included in Roger's bibliography[87] identifies some variety in techniques in research on change. Sample selection techniques

---

[87]E. Rogers, op. cit.

employed include random selection from a population and accidental selection through the use of intact groups. Since most of this research is historical or descriptive, there seems to be a scarcity of techniques for administering treatments. For the most part something happened to a group and it was studied ex post facto, usually with only an implied contrast group. Data collection seems most commonly to proceed through interview or survey and through participant observer techniques. Seldom is there a discussion of the validity or reliability of the obtained data. Descriptive statistics including per cent of response seem to be the most common data analysis. In some instances correlational and factor analysis are employed. This survey, admittedly restricted to Roger's bibliography, indicates that investigators of change have followed in a tradition of technique that is perhaps limiting real advancement in our understanding and control of change.

CRITERIA OF ADEQUACY FOR EVALUATING RESEARCH TECHNIQUES IN THE STUDY OF EDUCATIONAL CHANGE

The evaluation of techniques employed in a specific research project focuses upon the adequacy in conducting the specific techniques employed and the appropriateness of each technique in the research strategy. Time prohibits the enunciation of criteria for each possible technique. Generalized criteria on these techniques include those stated earlier in discussion of the design and data analysis research components. It seems important here to emphasize and expand on some of the already stated criteria.

Campbell and Stanley[88] present a generalized discussion of design in which two concerns are proposed, internal and external validity. Their suggestion focuses heavily on the experimental methodology. However, certain aspects have relevance to other methodologies.

Internal validity, that degree to which the study tests what it purports to test, depends upon the control of eight mediating variables.[89] These contributors to internal invalidity are

1. History--the specific events occuring between the first and second measurement or simply prior to a posttest in addition to the experimental variable.

2. Maturation--processes within the respondents operating as a function of the passage of time per se (not specific to the particular events), including growing older, growing hungrier, growing more tired, and the like.

3. Testing--the effects of having taken a pretest upon the scores of a second testing.

4. Instrumentation--changes in the calibrations of a measuring instrument or changes in the observers or scorers used which may produce changes in the obtained measurements.

---

[88] D. T. Campbell and J. C. Stanley, op. cit.

[89] Ibid. p. 171-245 para.

5. Statistical Regression--operating where groups have been selected on the basis of their extreme scores. The tendency of lower scorers on a test to score higher on a re-testing due to the presence of measurement errors. Also the tendency of high scorers to score lower on a retesting.

6. Selection - biases resulting from picking different respondents for the comparison groups.

7. Mortality - differential loss of respondents from the comparison groups.

8. Selection - maturation interactions, etc.-- where two of the previous factors working together might be mistaken for the effect of the experimental variable.

The sources of external invalidity are of equal concern as they restrict the applicability of the findings of a given study. Campbell and Stanley[90] suggest four such factors:

1. Interaction of testing and X--where a pretest might increase or decrease the respondents' sensitivity or responsiveness to the experimental variable and thus make the results obtained for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the experimental respondents were selected.

2. Interaction of Selection and X--the tendency of a typical subject to seek out or volunteer for a study, thus making the subjects different from persons in general.

---

[90] Ibid. p. 171-246 para.

3. Reactive Arrangements--effects of experimental arrangements which would preclude generalization about the effect of the experimental variable upon persons being exposed to it in nonexperimental settings.

4. Multiple Treatment Effects--in studies which involve alternating groups among treatments, one may come up with conclusions applicable only where this specific sequence of events is possible.

The relevance of these variables to methodologies other than experimental is demonstrated in the following examples. In a historical study, changes occurring within an institution may be maturational rather than the effect of some action. As a newly established school ages, the changing perceptions of each other on the part of the staff may be as important a factor in a change as is the action of a change agent. Instrumentaion changes affect both historical and descriptive methodologies. The historian who analyzes several documents may be classifying or evaluating the last document on the basis of different criteria than he did the first. The survey employing an interview may interpret the responses of the last subject in a different light than the first.

The works of Barker,[91] Rozenthal,[92] and the study of the

---

[91] R. G. Barker, The Stream of Behavior. New York: Appleton-Century-Crofts. 1963.

[92] R. Rozenthal, "Research on Experimenter Bias." Paper read at American Psychological Association, Cincinnati, September, 1959.

"Hawthorne Effect" all relate to the reactive arrangement category of external invalidity. When a person is engaged in an investigation, he tends to behave for the study in contrast to normal behavior. Orne[93] indicates that study is needed to determine what are the demand characteristics of a study. Barker[94] suggests a mode of study in which the investigator is not a participant but an observer in which the unaffected "stream of behavior" is recorded for analysis. The question vital here is, to what extent in a given study did the study of change affect change?

Another area of concern under the research technique rubric is the concern for sample selection techniques. The adequacy of a study of adoption of new staff utilization techniques in schools rest heavily on the extent to which the schools in the study represent schools in general. Cornell and McLoone[95] in reviewing the design of sample surveys indicate that in sampling attention must be paid to a precise description of the population or universe and to a determination

---

[93]Martin T. Orne, "On the Social Psychology of the Psychological Experience: With Particular Reference to Demand Characteristics and Their Implications," American Psychologist. 17:776-83; November 1962.

[94]R. G. Barker, "Explorations in Ecological Psychology." American Psychologist, 20:1-14; January 1965.

[95]F. G. Cornell and E. P. McLoone, "Design of Sample Surveys in Education." Review of Educational Research 33:523-32; December 1963.

of the permissible error and acceptable risk or confidence level to determine the appropriate sampling techniques.

The third area of research technique that needs evaluation is the data collection techniques. Here four concerns are expressed.

1. Is the data collection technique a valid measure?
2. Is the data collection technique a reliable measure?
3. What is the degree of objectivity of the data collection technique?
4. What is the practicality of the data collection technique?

Although these items are listed in relative importance, the researcher finds he must forfeit here to gain there. Thus, assessing adequacy involves the search for optimal conditions of validity, reliability, objectivity, and practicality.

The variety of data collection techniques is almost limitless. For example, the historian may examine records or documents or engage in interviews; the surveyer may observe, utilize standardized tests, projective techniques, interviews, etc. The experimentalist has equally as broad a variety. Research on change in the past has incorporated the participant observer, the accumulation and analysis of records, testing, s rveying, and interviewing.

The determination of the adequacy of a given technique requires the assessment of the appropriateness of the data so

collected for testing the hypothesis or answering the question. Secondly, adequacy rests on the assessment of the validity-reliability-objectivity-practicality of the measuring device.

In assessing instrument adequacy of the evaluation of research, one must (1) understand the types of validity and reliability and (2) interpret the importance of these for a given study. Under the rubric validity four categories exist.

1. Content validity--the degree to which the items in the measuring device are contained in treatment of which it is a measure.

2. Construct validity--the degree to which the instrument accumulates evidence on some hypothetical construct.

3. Concurrent validity--the degree to which the instrument accumulates data which correlated with concommitant performance.

4. Predictive validity--the degree to which the instrument accumulates data which correlates with performance at some future date.

It is clear that, although we should be concerned with each of the above, for a given study one may be more important than another.

Our concern for the reliability of measuring devices takes three forms:

1. Internal consistency--the degree to which all parts of an instrument are measuring the same thing.

2. Stability--the degree to which subsequent administrations of the instrument accumulate comparable data.

3. Equivalence--the degree to which alternate forms of an instrument accumulate comparable data.

Again, although each is a vital concern, the relative importance of these items is determined by the specifics in a given study.

The failure of a researcher to attend to both of these areas in his study and to communicate this in his report is severely depricating to the adequacy of the study. Without information of this sort we know not what we have measured nor to what extent we would obtain the measurement again.

The fourth and final area of concern regarding research techniques is the area of data analysis. Through the literature admonition can be found that appropriate and modern analysis techniques must be used. The definition of "appropriate" can be found in the study of assumptions inherent in statistical models. "Modern," however, is an undefinable term and thus its use as a criterion is difficult. It is assumed that in any given report the analytical technique employed was as modern as possible for the investigator. In light of newer analytical techniques, the evaluator of research may infer weakness in any given study.

The appropriateness of a statistic is based upon (1) the scalar characteristics of the data, (2) the number of variables, (3) the relationship among the variables, and (4) the manner in which randomization is inserted in the data generation. Senders[96]

---

[96] V. L. Senders, op. cit.

presents a clear description of the four categories of scales, nominal, ordinal, interval, and ratio. An argument rages as to whether statistics based upon interval assumptions can be used with ordinal data. It can be demonstrated that some kinds of knowledge can be gained through averaging ordinal data. Perhaps a continuum of scalar quality exists. The crucial point seems not to be the employment of the incorrect statistic but rather the interpretation of the finding. The use of any mathematical process on frequency data presents findings regarding frequency of responses rather than quality of response.[97]

The determination of the appropriate inferential statistic depends upon the number and the degree of independence of the variables. This is demonstrated in grid form by the presentations by Senders,[98] Tatsuoka and Tiedeman,[99] and Siegel.[100]

CONCLUSION

Several criteria for research adequacy, an attempt at definition of the field of research on educational change, and

---

[97] W. L. Hays, Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963. p. 73.

[98] V. L. Senders, op. cit. p. 256-7.

[99] M. M Tatsuoka and D. V. Tiedeman, op. cit. p. 145-5.

[100] S. Siegel, op. cit. (Inside Cover)

specific criteria for research techniques have been presented. Two points must be made in closing. To date, standards have not been drawn by which to measure the adequacy of a specific problem statement, hypothesis, etc. The field seems to have settled on what must be done, and through this, apparently, agreement on research evaluation can be reached. Despite this lack of a final yardstick, research adequacy must be assessed if progress is to be made in the accumulation of knowledge or if the field is to be sure of that which is known.

The second point relates to the general negativism that evolves from a systematic analysis of research. The Encyclopedia of Retrospect is a powerful book. Through hindsight we can observe all manner of error unobservable to foresight. However, without the action based upon limited foresight, hindsight is severly reduced. It is here proposed that Professor Designbumbler did not set out to conduct a fallacious study. He did the best he could with the materials at hand and his level of sophistication. Rather than berating his competence personally, the field will progress if the focus is on what was done right and wrong, what was left undone, and what can be done to build on these.

BIBLIOGRAPHY

American Institute of Research, "A Procedure for Evaluating Graduate Research on the Basis of the Thesis." Pittsburg: October 1955.

Backrack, A. J., Psychological Research: An Introduction. New York: Random House, Inc., 1962.

Barker, R. B., The Stream of Behavior. New York: Appleton-Century-Crofts, 1963.

Barker, R. B., "Explorations in Ecological Psychology." American Psychologist. 20:1-14; January, 1965.

Barr, Arvil S., Robert A Davis, and Palmer O. Johnson, Educational Research and Appraisal. Chicago: J. E. Lippincott Company, 1953.

Best, John W., Research in Education. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1959.

Bixler, H. H., "Checklists for Educational Research." New York: Teachers College, Columbia University, 1928. p. 85-7.

Borg, Walter R., Educational Research: An Introduction. New York: David McKay Company, Inc., 1963.

Campbell, Donald T. and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," in Handbook of Research on Teaching. N. L. Gage, Editor, Chicago: Rand McNally, 1963.

Clark, David L., Egon G. Guba, and Gerald R. Smith, "Functions and Definitions of Functions of a Research Proposal or Research Report." Unpublished mimeo, Columbus, Ohio: The Ohio State University, 1962.

Cyphert, Frederick R., and Ernest Spaights, "An Analysis and Projection of Research in Teacher Education," Cooperative Research Project No. F-015; The Ohio State University Research Foundation, 1964.

Dooley, Mother M. Constance, "The Relationship Between Arithmetic Research and the Content of Arithmetic Textbooks, (1900-1957)" The Arithmetic Teacher 7:178; April, 1960.

Farquahar, William W. and John D. Krumboltz, "A Checklist for Evaluating Experimental Research in Psychology and Education." Journal of Educational Research 52:534-5.

Gephart, W. J., "Development of an Instrument for Evaluating Reports of Educational Research." Washington, D. C.: Cooperative Research Project S-014, 1964.

Good, Carter B., Introduction to Educational Research: Second Edition. New York: Appleton-Century-Crofts, Inc., 1954.

Guba, Egon G., "The Writing of Proposals" in Research in Educational Administration. Stephen P. Hencley, Editor, Comparative Research Project No. F-2, Washington, D. C., January 1962.

Hart, R. N., "What Does the Elementary School Counselor Do?" The School Counselor 9:70-2; December 1961.

Hays, William L., Statistics for Psychologists. New York: Holt, Rineholt and Winston, 1963.

Hillway, Tyrus, Introduction to Research: Second Edition. Boston: Houghton Mifflin Company, 1964.

Johnson, Granville, B., Jr., "A Method for Evaluating Research Articles in Education." Journal of Educational Research 51:149-51; October, 1957.

Kerlinger, F. N., Foundations of Behavioral Research New York: Holt, Rinehart and Winston, Inc., 1964.

King, W. B., Survey of the Status of Research in Guidance and Counseling. Washington, D. C.: U.S.O.E. Cooperative Research Project Number F-1, 1962.

Lake, D. E., "Inductive Methodology Versus Hypothetic-Deductive Methodology in Educational Research." Unpublished doctoral dissertation, University of Kansas, 1961.

Lamke, T. A.,"Primer in Research: Lesson I. Identifying and Defining the Problem." Phi Dalta Kappan 38:127-33; January 1957.

Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Company, 1953.

Lindvall, C. M., "Review of Related Research." Phi Delta Kappan 40:179-80; January 1959.

McGrath, G. D., James J. Jelinek, Raymond E. Wochner, Educational Research Methods. New York: The Ronald Press Company, 1963.

McNeil, J. D., "Programed Instruction Versus Usual Classroom Procedures in Teaching Boys to Read." American Educational Research Journal 1:113-20; March 1964.

McNemar, Quinn, Psychological Statistics. New York: John Wiley and Sons, Inc., 1955.

Monroe, Walter S. and M. D. Engelhart, The Scientific Study of Educational Problems. New York: The Macmillan Company, 1936.

Mouly, George J., The Science of Educational Research. New York: American Book Company, 1963.

Orne, Martin T., "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications." American Psychologist 17: 776-83; November 1962.

Platt, J. R. "Strong Inference," Science 146:347-52; October 1964.

Rogers, E., Diffusion of Innovations. New York: The Free Press of Glencoe, 1962.

Senders, Virginia, Measurement and Statistics. New York: Oxford University Press, 1958. p. 524-7.

Siegel, Sidney, Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill Book Company, Inc., 1956. Inside Cover.

Smith, G. R., "Inadequacies in a Selected Sample of Research Proposals." Unpublished doctoral dissertation. Teachers College, Columbia University, 1964.

Stanley, J. C., "Improvement of Educational Research." Presented to the Seventh Annual Phi Delta Kappa Research Symposium. Madison, Wisconsin, 1965.

Symonds, Percival, "A Research Checklist in Educational Psychology." Journal of Educational Psychology 47:100-109; February 1956.

Tatsuoka, M. M. and David B. Tiedeman, "Statistics as an Agent of Scientific Method in Research on Teaching," in Handbook of Research on Teaching. N. L. Gage, Editor; Chicago: Rand McNally and Company, 1963.

Travers, Robert M. W., An Introduction to Educational Research: Second Edition. New York: The Macmillan Company, 1964.

Van Dalen, D. B., Understanding Educational Research. New York: McGraw-Hill Book Company; 1962.

Van Dalen, D. B., "Research Checklist in Education." Educational Administration and Supervision 44:174-81, May 1958.

Walton, R. E., 'Two Strategies of Social Change and Their Dilemmas," The Journal of Applied Behavioral Sciences 1:167-79; April-May-June 1965.

Whitney, Frederick L., The Elements of Research: Revised Edition. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1942.

Wilson, G. M., "Research: Suggested Standards for Summarizing and Reporting Applied to Two Recent Summaries of Studies in Arithmetic." Journal of Educational Research 28:187-94; November, 1934.